PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/11, C12P 21/08, C12Q 1/68, C12N 15/62, C07K 15/28

A2

(11) International Publication Number: WO 94/01548

(43) International Publication Date: 20 January 1994 (20.01.94)

(21) International Application Number: PCT/GB93/01467

(22) International Filing Date: 13 July 1993 (13.07.93)

(30) Priority data:
9214857.6 13 July 1992 (13.07.92) GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)* : SIBSON, David, Ross [GB/GB]; 37 Grimsdells Lane, Amersham, Buckinghamshire HP6 6HF (GB). GROSS, Jacqueline [GB/GB]; 47 Boxmoor Road, Kenton, Middlesex HA3 8LH (GB). HADFIELD, Kathryn, Mary [GB/GB]; 5 Carlisle Terrace, St Ives, Huntingdon, Cambridgeshire PE17 4PQ (GB). HOWELLS, David [GB/GB]; 77 Puttocks Drive, Welham Green, Hatfield, Hertfordshire AL9 7LW (GB). STARKEY, Michael [GB/GB]; 27 Creasy Close, Abbots Langley, Hertfordshire WO5 0HS (GB). KELLY, Maria [IE/GB]; 24A Oxford Road, Ealing, London W5 3ST (GB). SHAW, Diana [GB/CA]; 342 Glacier Hall, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4 (CA).

(74) Agent: BIZLEY, Richard, Edward; Hepworth Lawrence Bryer & Bizley, 2nd Floor Gate House South, West Gate, Harlow, Essex CM20 1JN (GB).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE, PLACENTA OR BONE NARROW

(57) Abstract

This invention provides a nucleic acid fragment encoding a gene product or portion thereof and comprising any one of: (a) a sequence selected from SEQ ID Nos 1 to 1193 from the attached sequence listings; (b) an allelic variation of a sequence as defined in (a); or (c) a sequence complementary to (a) or (b). The invention includes uses of such fragments, and gene products corresponding thereto.

HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE, PLACENTA OR BONE NARROW

This invention relates to new nucleic acid fragments encoding gene products or portions thereof, which fragments are obtainable from human nucleic acid populations, individual members of such populations being present in widely varying amounts.

Situations are increasingly arising in which it is necessary to study complex nucleic acid or polynucleotide populations. For example, it is now widely appreciated that an invaluable resource could be created if the entire sequence of the genomes of organisms such as man were determined and the information available. The magnitude of such a task should not, however, be underestimated. Thus, the human genome may contain as many as 100,000 genes [a very substantial proportion of which may be expressed in the human brain (Sutcliffe, Ann. Rev. Neurosci. 11:157 (1988))]. Only a very small percentage of the stock of human genes has presently been explored, and this largely in a piecemeal and usually specifically targeted fashion.

There has been much public debate about the best means of approaching human genome sequencing. Brenner has argued (CIBA Foundation Symposium 149:6 (1990)) that efforts should be concentrated on cDNAs produced from reverse transcribed mRNAs rather than on genomic DNA. This is primarily because most useful genetic information resides in the fraction of the genome which corresponds to mRNA, and this fraction is a very small part of the total (5% or less). Moreover, techniques for generating cDNAs are also well known. On the other hand, even supposing near perfect recovery of cDNAs corresponding to all expressed mRNAs, some potentially useful information will be lost by the cDNA approach, including sequences responsible for control and regulation of genes. Nonetheless, the cDNA approach at least substantially reduces the inherent inefficiencies resulting from analysis of repeated sequences or non-coding sequences in an approach which depends upon genomic DNA sequencing.

Recently, the results of a rapid method for identifying and characterising new cDNAs has been reported (Adams, M.D. et al., Science 252, 1991, pp 1651-1656). Essentially, a semi-automated sequence reader was used to produce a single read of sequence from one end of each of a number of cDNAs picked at random. It was shown, by comparing the nucleic acid sequences of the cDNAs (or the protein sequences produced by translating the nucleic acid sequence of the cDNAs) to each

A third approach exploits the second order reassociation kinetics of cDNA annealing to itself. After a long period of annealing, the cDNAs which remain single stranded will have nearly the same abundance, and can be recovered by standard PCR (see Patanjali, S.R. et al., PNAS USA 88, 1991, pp. 1943-1947; Ko, M.S.H., NAR 19, no.18, 1991, pp 5705-5711). The methods disclosed in these two publications, however, suffer from notable disadvantages. They are entirely dependent on the stringent physical separation of single stranded and double stranded DNA, require an elevated number of manual manipulations in each reaction, and necessitate protracted hybridisation times (up to 288 hours in the method of Patanjali et al.)

Yet a further approach in "normalising" a nucleotide population is described in co-pending British Patent Application No 91 15407.0, filed 17th July, 1991 by MRC, and involves a PCR process in which a mixture comprising a heterogenous DNA population and appropriate oligonucleotide primers is first formed and the DNA denatured, but before effecting a conventional PCR protocol the conditions are altered to allow the denatured strands of the more common DNA species to reanneal together, whilst avoiding annealing of primers to the DNA strands. By this means, rarer species can subsequently be amplified in preference to the more common species.

This PCR normalisation method in general comprises the steps of:

(a) preparing a mixture comprising a heterogenous DNA population and oligonucleotide primers suitable for use in a PCR process, in which the DNA is denatured;

(b) altering the conditions to allow the denatured strands of the more common DNA species to reanneal, while preventing the annealing to the primers to the DNA strands;

(c) further altering the conditions of the mixture in order to allow the primers to anneal to the remaining single-stranded DNA comprising the rarer DNA species; and

(d) carrying out an extension synthesis in the mixture produced in step (c).

Advantageously, the method consists of a cyclic application of the above four steps.

preferably an endonuclease which has its cleavage and recognition sites separated, which reagent cleaves the nucleic acid so as to produce double stranded cleavage products the individual strands of which overlap at cleaved ends to leave a single strand extending to a known extent;

(b) ligating the cleavage products from (a) with a population of adaptor molecules to generate adaptored cleavage products, each of which adaptor molecules has a cleavage product end recognition sequence and the population thereof encompassing a range of adaptor molecules having recognition sequences complementary to a predetermined subset of the sequences of the cleavage-generated extending single strands; and

(c) selecting and separating only those adaptored cleavage products resulting from (b) which carry an adaptor of predetermined recognition sequence.

A preferred endonuclease for use in step (a) of the above process is Fok 1.

An important feature of this process is the use of adaptor molecules. The adaptors used must have "overhanging" fragment recognition sequences which reflect or are complementary to the extending cleavage-derived sequences which the adaptors are designed to react with. It is also preferred that the adaptors used should end with a 5' hydroxyl group. The avoidance of a 5' phosphate group removes the risk of inappropriate ligation involving the adaptors.

Adaptor molecules may also contain a portion permitting specific sequence selection and separation (as in step (c) of the process) when a sequence is attached to the adaptor. For example, an adaptor can carry biotin, thereby permitting advantage to be taken of the biotin/avidin
reaction in selecting and separating desired adaptored molecules. Additionally, adaptors preferably comprise a known and selected sequence such that specifically isolated adaptored molecules can be amplified by known techniques (such as PCR) using a primer complementary to the core sequence.

Preferably the adaptors are short double-stranded oligonucleotides which can be joined to the ends of cleavage products. They will have been chemically synthesised so that their sequence can be predetermined

chosen, e.g. for use in subsequent nucleotide sequencing. This facilitates, for example, the identification of a large population of sequences by permitting a rational approach to splitting such populations into subsets, each of which subsets can be examined in turn.

In the light of these developments, the present invention now provides a nucleic acid fragment encoding a gene product or portion thereof and comprising any one of:-

(a) a sequence selected from SEQ ID Nos 1 to 1193;

(b) an allelic variation of a sequence as defined in (a); or

(c) a sequence complementary to (a) or (b).

In another aspect, the invention provides a nucleic acid sequence as set out in any one of SEQ ID Nos 1 to 1193, or a complement or allelic variation thereof. Preferred sequences exhibit no more than 90% homology to a human sequence known per se.

In a further aspect, the invention provides a nucleic acid fragment comprising a portion of a sequence as defined above of sufficient size such that a probe of the same size and exhibiting complementarity to said portion can hybridise to said sequence. Preferably, such portions are at least 15 bases in length. It will be appreciated that minor mismatches in the aforesaid "complementarity" are not excluded provided hybridisation can still occur. In general, hybridisation conditions are within the choice of the skilled person, but reference can be made, for example, to the following: Melting temperature of hybrids - Bolton, E. T. and McCarthy, B. J. Proc. Natl. Acad. Sci, 48 p1390 (1962). Effect of formamide on lowering melting temperature - Casey, J. and Davidson, N., Nucleic Acids Res. 4, p1539 (1977). Effect of imperfect homology - Bonner, T. I. et al., J. Mol. Biol. 81, p123 (1973). General - Meinkoth, J. and Wahl, G. Anal. Biochem. 138, p267 (1984). Oligo hybridization and washing - Lathe, R. J. Mol. Biol. 183, P1 (1985).

The present invention also envisages DNA constructs comprising fragments or sequences as referred to above with a control or regulatory sequence.

D. M. et al., p108 Blackwell scientific Publications, Oxford. Harlow, E and Lane, D. Antibodies, A Laboratory Manual, Cold spring harbor Laboratory, Cold Spring, New York.

Expression in an appropriate higher eucaryotic host may be important to ensure correct protein folding and also activity. Expression to avoid copurification of toxic products can sometimes be better performed in organisms approved for human consumption, eg prokaryotic Bacillus subtilis, eurkaryotic yeast, mammalian cows milk vectors, and other methods known in the art.

The invention also includes novel gene products or portions thereof encoded by a fragment, sequence or gene-comprising DNA fragment of the invention.

It will be appreciated that the sequences of the present invention collectively have utility based, inter alia, upon their common origin, and hence they can effectively be considered together rather than as separate entities. It is convenient to represent them as separate sequences, because this is how they were produced and serves as "punctuation" between the different functional entities which each sequence represents. However, the sequences could just as easily have been presented as a continuous sequence derived by placing them end to end in the order in which they were produced, with a separate indication of where the beginnings and ends of the component sequences are.

In contrast to investigations hitherto, where gene fragments (sequence fragments) could only be identified through some known characteristic [for example: their homology to a fragment which largely encodes amino acids identified by sequencing a previously isolated peptide or is the antisense of that coding sequence; or them having at least partial homology to previously characterised nucleic acids; or them having ability to encode expressed proteins which could later be detected by functional assays of the cells expressing those proteins or by using antibodies which had been previously raised against the proteins to detect their expression, Sambrook J., et al., Molecular Cloning CSH Press 1989], the sequences and fragments described by the present invention are entirely underivable and unpredictable from the prior art, but are nonetheless clearly of great value for various purposes.

Thus, such sequences, by comparing them to sequence databases, can be

eucaryotes may be preferred to ensure correct modifications, for example, glycosylation in vivo, when this proves to be important. Expression can be ensured by situating the cDNA appropriately to signals for expression (Amann, E. and Brosius, J. Gene 40 p183 1985), Shimuzu, Y et al., Gene 65, p141 (1988), Straus, D. and Gilbert, W. Proc. Natl. Acad. Sci. 82, p2014 (1985)). Such signals may include a promoter for transcription, which may itself be regulatable.

The proteins thus-expressed can be screened for activities of therapeutic or commercial value. It may be that the proteins have to be first isolated for this purpose or can be assayed in situ. It may be desirable that some means of stabilising the expressed protein is employed. This can be achieved, for example, (and as indicated earlier) by expressing in frame as part of a fusion polypeptide (Smith, D. B., et al., Proc. Natl. Acad. Sci. 83 p8073 (1986)).

Useful antibodies can be raised against the expressed proteins. It is commonly not an absolute requirement that full length proteins are produced, although this may influence the quality of the antibodies produced. Peptides as short as 8 or 9 amino-acids in length can be used as antigens (Germain R., N. Nature 353 pp605-607 (1991), Rudensky, A., Y., et al., Nature 353 p622-627 (1991)). Immunogenic peptides could simply be synthesised using the amino-acid sequence translated from a sequence or fragment of this invention. It is desirable, although not absolutely required, that some means of producing purified antibodies is adopted. When fusion polypeptides are used to raise antibodies, an affinity matrix specific for the generic part of the protein allows the fusion polypeptide to be immobilised (Smith, D. B., et al., Proc. Natl. Acad. Sci. 83 p8073 (1986)). The immobilised polypeptide can then be used to affinity purify the antibodies. Antibodies to both the generic part of the fusion polypeptide and the part of interest are produced. When these need to be discriminated between, a different affinity column can be used to remove only those antibodies specific for the generic part of the polypeptide. Alternatively, and as mentioned earlier, it can be arranged that the boundary between the two separate protein components of the fusion polypeptide has the recognition sequence for an endopeptidase with a rare cutting site. The peptide of interest can then be released from the affinity purified polypeptide by the action of the endopeptidase (Nagai, K., and Thogersen, H., C. Methods Enzmol. 153 p461-481 (1987). Another alternative is raise monoclonal antibodies against the purified protein.

clone.

This invention will now be further described and illustrated by means of the following Examples.

All oligonucleotides used in these Examples were synthesised Trityl on using an ABI 380B DNA Synthesizer according to the manufacturers instructions. Purification was by reverse phase HPLC (see, for example, Becker, C., R., et al., J. Chromatography 326, p293-299 (1985)).

Example 1

Human brain and adrenal tissues were obtained from a mixture of 12 to 15 week menstrual age foetuses and then snap frozen in liquid nitrogen before storing in bijou bottles in a -80°C freezer. The two types of tissue were used separately, directly from the freezer, to prepare cDNA from which restriction fragments were generated for sorting into subsets. 1g portions of each of the separate tissues were homogenised, using an Ultra-Turrax T25 Disperser (Janke and Kunkel, IKA-Labortechnik), on ice in the presence of 4M guanidinium isothiocyanate to solubilise macromolecules. RNA was isolated from each homogenate by using centrifugation to sediment it through caesium trifluoroacetate. This was performed using the Pharmacia kit according to the manufacturer's instructions, except that centrifugation was performed for 36 hours and the RNA obtained was finally desalted and concentrated by performing two ethanol precipitations in succession with two 70% ethanol washes after each precipitation. In each case, polyA$^+$ (mRNA) was isolated from 200 to 400 μg of the total RNA by binding it to magnetic oligo-dT coated beads (Dynal). Solution containing unbound material was removed from the beads, which were washed, and then mRNA eluted directly for use. mRNA isolation was performed in accordance with the manufacturer's instructions. Yields of RNA from the beads were between 1 and 3% of the total RNA. 2 to 4 μg of the eluted RNA were used for cDNA synthesis. cDNA synthesis was performed according to the method of Gubler, U and Hoffman, (B. J. Gene 25 p263 (1983) using a Pharmacia kit according to the manufacturer's instructions. OligodT was used to prime the first strand cDNA synthesis reaction. The cDNA was purified by extracting twice with phenol/chloroform and then low molecular weight solutes including nucleic acids below ca. 300 bases were removed by passing the cDNA reaction mixture through a Pharmacia S400 spun column used according to the manufacturer's

GTTCTCGGAGCACTGTCCGAGA (SEQ ID: 1198) biotinylated adaptor. Bead binding was allowed to proceed at 28°C for 30 minutes with mixing every 10 minutes.

Un-biotinylated cDNAs were washed from the beads with 400μl each of 2M NaCl twice, fresh 0.15 mM NaOH four times at 28°C for 5 minutes each, water twice and finally a buffer comprising 20 mM Tris pH 8.3, 50 mM NaCl, and 25mM Mg^{2+}. The beads were then resuspended in 240 μl of the final buffer including additionally 0.5 mM dNTPs and divided into 4x60 μl.

Four of the 60 μl aliquots, two from each tissue, were processed further specifically to prime and copy a subset of the immobilised, adaptored fragments. 2 pmoles of the primer 5' CTGTCTGTCGCAGGAGAAGGAA (SEQ ID: 1201) were added to each of two aliquots, one from each tissue. 2 pmoles of the primer 5' CTGTCTGTCGCAGGAGAAGGAG (SEQ ID: 1202) were added to each of the other two aliquots. 2.5 units of Taq DNA polymerase were added to each reaction and 16 cycles of alternate denaturation at 95°C for 30 seconds, annealing at 63°C for 2 minutes and polymerisation at 72°C for 3 minutes was performed to accumulate the selected single-strands in solution.

On completion of the DNA synthesis reactions a further 30 μl of resuspended beads were added to each reaction to remove the biotinylated fragments. The reaction was incubated at 28°C for 30 minutes mixing every 10 minutes to ensure that the biotinylated strands were bead bound. Each aqueous phase containing the newly synthesised strands was then removed and extracted with phenol/chloroform twice to remove the enzyme before being further purified by passing through an S400 spun column equilibrated with 10 mM Tris pH 8.3/50 mM NaCl as described above.

Rounds of PCR amplification of subsets of the selected fragments were performed by using the original primer in each case, together with one of the primers 5' GTTCTCGGAGCACTGTCCGAGAG (SEQ ID: 1199) or 5' GTTCTCGGAGCACTGTCCGAGAC SEQ ID: 1200). This simultaneously rendered the fragments double-stranded and increased the amounts of available material. It was not known how many cycles of amplification would be required at this stage, since each primer pair would be expected to behave differently. It was therefore necessary directly to determine a suitable number empirically by using standard agarose gel electrophoresis to examine the reaction products after a given number

and Short J. M., Nucl. Acids Res. 17 p9494) was prepared for accepting the resected cDNA by restriction cleavage at the BamHI and HindIII sites and then adaptoring the resultant cohesive ends using the specific adaptors produced by the oligonucleotide 5' AGCTCGGCTCGAGTCTG (SEQ ID: 1203) with its partially complementary oligonucleotide 5' GCGACAGACAGCAGACTCGAGCCG (SEQ ID: 1204) and the oligonucleotide 5' GATCCGGCTCGAGT (SEQ ID: 1205) with its partially complementary oligonucleotide 5' CCGAGAACACTCGAGCCG (SEQ ID: 1206). Preparation of the vector and adaptoring were performed according to standard procedures. Insertion of the cDNA was performed between the BamHI and HindIII restriction sites. Recombinant vectors were transformed into the host XL1-Blue (Bullock, W. O. et al Biotechniques 5 p376-378 (1987)) by the method of Hannahan, D. J. (Mol. Biol. 166 p577-580 (1983)). Suitable standard controls for the ligations and transformations were also included.

Post transformation procedures were as described in "Molecular Cloning", 2nd Edition (Sambrook J., Fritsch, E. F., and Maniatis, T. CSH Press (1989)). Colonies were produced by plating onto X-gal/IPTG L-agar plates containing 50μg/ml ampicillin and 10μg/ml tetracyclin. Clear colonies were picked, each into a separate well of a microtitre plate, containing 100μl of L-broth and 50μg/ml ampicillin. Growth was allowed to occur for 16 hours at 37°C. 100μl of 50% or 30% glycerol was added to plates which were archived at -20°C or -80°C, respectively.

Bacteria corresponding to those archived were used for preparing templates for sequencing by the dideoxy method (Sanger, F. Milklen, S. and Coulson, A. R. Proc. Natl. Acad. Sci. 74 p5463-5467 (1977)). Bacteria for this purpose were either grown on L-agar plates containing 50μg/ml of ampicillin, prepared at the same time as they had been grown in liquid culture, or after plating out from the archive. Alternatively, fresh liquid cultures were inoculated from the archive. In all cases, cDNA inserts were amplified for sequencing by PCR (Saiki, R. K. et al Science 239 p487-491 (1988)). PCR was either performed using bacteria directly added to the reaction, by a toothpick, or PCR was performed using 1/50th of the plasmid isolated by preparative methods (Holmes, D. S. and Quigley, M. Anal. Biochem. 114 p193 (1981)) from the bacteria in the liquid cultures or from the plates.

20 pmoles of each of the PCR primers 5' biotinylated GTAAAACGACGGCCAGT

A second method of preparing cDNA libraries for obtaining gene fragments of the invention took advantage of the PCR normalisation process described above. Standard procedures were used to prepare mRNA from RNA that had been isolated by standard caesium chloride bouyant density gradient methods from a full term human placenta. The oligonucleotide LNotdt, sequence 5' TACGTTCGACAAGCTTGAATTCGCGGCCGC$(T)_{26}$, (SEQ ID: 1210) was used at 1 μM with AMV reverse transcriptase, to prime first strand cDNA synthesis under standard conditions from 0.5 μg of the placental mRNA. Temperatures above 65°C were used to inactivate the reverse transcriptase and then the volume of the reaction made up to 100 μl with water.

PCRs were then performed in reactions containing 1 μl of the diluted cDNA, 10 mM Tris-HCl pH 8.3, 40 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 μM dNTPs, 10 uCi $a^{32}P$ dCTP, 1 μM each of the primers 11AD1, sequence 5' GCC(TA)(GC)CGCCGA (SEQ ID: 1211), and LNotdT and Taq DNA polymerase. An initial denaturation period of 95°c for 90 seconds was followed either by 35 cycles of standard PCR, comprising 95°C for 30 seconds, 45°C for 30 seconds and 72°C for 30 seconds or alternatively 3 cycles of the standard PCR already described followed by 27 cycles of Cot PCR during which an additional step of 72°C for 16 minutes was placed between all of the 95°C and 45°C steps of the standard PCR. The standard PCR was followed by a single 72°C for 3 minutes step while the Cot PCR was followed by one standard PCR cycle except that the 72°C incubation was performed for 3 minutes.

Products of the PCR reaction were end repaired by adding 5 units of T4 DNA polymerase to the reaction and then incubating at 37°C for 10 minutes. Enzymes were removed by phenol extraction. The cDNA was precipitated by 70% ethanol, dried and then resuspended in NotI buffer. 20 units of NotI were used to digest the cDNA under standard conditions. cDNA was again phenol extracted and ethanol precipitated. 10% of the purified NotI cut DNA were ligated to the vector pBluescript lting-Meese, M. A. and Short J. M. Nucl. Acid Res. 17 p9494 which had been prepared as standard to receive this DNA by restricting with the enzymes NotI and EcoRV. Transformation and processing of clear colonies was performed as described above except that the host E.coli strain DH5a was used in place of XL-1 Blue.

Preparation of clones for sequencing, sequencing and sequence analysis of cDNAs in clones thus-produced were performed as described in Example 1.

CLAIMS:

1. A nucleic acid fragment encoding a gene product or portion thereof and comprising any one of:-

(a) a sequence selected from SEQ ID Nos 1 to 1193;

(b) an allelic variation of a sequence as defined in (a); or

(c) a sequence complementary to (a) or (b).

2. A nucleic acid sequence as set out in any one of SEQ ID Nos 1 to 1193, or a complement or allelic variation thereof.

3. A sequence as claimed in claim 2 and which exhibits no more than 90% homology to a human sequence known per se.

4. A nucleic acid fragment comprising a portion of a sequence as defined in claim 2 or claim 3 of sufficient size such that a probe of the same size and exhibiting complementarity to said portion can hybridize to said sequence as defined in claim 2 or claim 3.

5. A fragment as claimed in claim 4, wherein said portion is at least 15 bases in length.

6. A fragment as claimed in any one of claims 1, 4 or 5 and encoding at least a portion of a biologically active polypeptide.

7. A nucleic acid sequence as claimed in claim 2 or claim 3 and encoding at least a portion of a biologically active polypeptide.

8. A DNA construct comprising a fragment as defined in any one of claims 1, 4, 5 or 6 or a sequence as defined in any one of claims 2, 3 or 7, together with a control or regulatory sequence.

9. A construct as claimed in claim 8 which encodes a fusion protein comprising a known protein and the polypeptide encoded by said fragment or sequence.

10. A construct as claimed in claim 9, wherein the fusion protein encoded is a cleavable fusion protein having an endopeptidase recognition site positioned between codons corresponding to said known

protein and said fragment or sequence.

11. The use of a fragment as defined in any one of claims 1, 4, 5 or 6 or a sequence as defined in any one of claims 2, 3 or 7 to produce a gene.

12. A DNA fragment comprising a gene obtainable by the use defined in claim 11.

13. An expression vector comprising a fragment as defined in any one of claims 1, 3, 5 or 6, a sequence as defined in any one of claims 2, 3 or 7, a DNA construct as defined in any one of claims 8 to 10, or a DNA fragment as claimed in claim 12, positioned such that that nucleic acid sequence which encodes the polypeptide corresponding to said fragment, sequence or DNA fragment is in operable reading frame with a control or regulatory sequence.

14. A vector as claimed in claim 13, wherein said vector control or regulatory sequence comprises a regulatable promoter.

15. Host cells which incorporate as a heterologous part of their expressible genetic information a fragment as defined in any one of claims 1, 3, 5 or 6, a sequence as defined in any one of claims 2, 3 or 7, or a DNA fragment as defined in claim 12.

16. A process for the production of a polypeptide comprising cultivating host cells as defined in claim 15.

17. An antibody directed against a polypeptide obtainable by the performance of a process as defined in claim 16.

18. An antibody as claimed in claim 17 and which is monoclonal.

19. A novel gene product or portion thereof encoded by a fragment as defined in any one of claims 1, 3, 5 or 6, or encoded by a sequence as defined in any one of claims 2, 3 or 7, or encoded by the gene comprised in a DNA fragment as defined in claim 12.